League tables and school effectiveness: a mathematical model

Rebecca B. Hoyle^{1*} and James C. Robinson²

¹*Department of Mathematics and Statistics, University of Surrey, Guildford, Surrey GU2 7XH, UK*

²*Mathematics Institute, University of Warwick, Coventry CV4 7AL, UK*

'School performance tables', an alphabetical list of secondary schools along with aggregates of their pupils' performances in national tests, have been published in the UK since 1992. Inevitably, the media have responded by publishing ranked 'league tables'. Despite concern over the potentially divisive effect of such tables, the current government has continued to publish this information in the same form. The effect of this information on standards and on the social make-up of the community has been keenly debated. Since there is no control group available that would allow us to investigate this issue directly, we present here a simple mathematical model. Our results indicate that, while random fluctuations from year to year can cause large distortions in the league-table positions, some schools still establish themselves as 'desirable'. To our surprise, we found that 'value-added' tables were no more accurate than tables based on raw exam scores, while a different method of drawing up the tables, in which exam results are averaged over a period of time, appears to give a much more reliable measure of school performance.

Keywords: league tables; school effectiveness; value-added measures

1. INTRODUCTION

The 'Parent's Charter', first published by the British Conservative government in 1991, promised the publication of examination results to allow parents to exercise 'informed' choice over their children's schooling. A basis for national testing and the gathering of comparative data was provided by the Education Reform Act (UK Government Department of Education and Science 1988), which established a uniform National Curriculum.

The resulting 'school performance tables' for secondary schools in England, Scotland and Wales were first published in 1992 (Department for Education 1992) under the then Secretary of State for Education, John Patten. They take the form of an alphabetical list of schools with details of their public examination results. Primary school tables followed in 1997 under Gillian Shephard, and are based on the performances of 11 year olds in Key Stage 2 tests (standardized national tests introduced by the Department for Education and Employment 1996).

The media quickly seized upon this information, publishing 'league tables' that rank schools from best to worst according to their examination results.

School league tables are not unique to the UK: standardized tests and league tables on the British model are now being piloted in Italy under the auspices of the Istituto nazionale per la valutazione del sistema dell'istruzione (Newbold 2000); Australia publishes data on individual schools' exam results; and 45 states in the USA publish 'report cards' on schools, out of which 27 rate schools or identify low-performing schools (*Education Week* 2001).

League tables are becoming increasingly common beyond the school sector as governments seek to provide greater information to the public across a wide range of services. University and MBA league tables are now com-

monplace in the British and American press; in the UK this trend has accelerated in response to the ready availability of 'quantitative' data from the Research Assessment Exercise and Teaching Quality Assessments. Hospitals and healthcare providers are also ranked in tables published in both the UK and the USA. In the UK the Patient's Charter led to the publication of 'high level indicators' and 'clinical indicators' for all National Health Service hospitals in June 1999 (NHS Executive 1999*a,b*) and the introduction of controversial 'star' ratings for hospitals in September 2001 (Department of Health 2001).

Whilst the aim of bringing greater transparency to our public services is admirable, the effectiveness of publishing the data in this form is not entirely clear. School league tables have been widely seen as divisive and misleading by parents, teachers and many educational researchers, but none the less have been continued in the UK by the current Labour government. However, two of the regions with their own legislative assemblies devolved from central Westminster control chose to scrap the tables in 2001: first Northern Ireland (where league tables were first published in 1993) and then Wales.

There are two main arguments against school league tables as currently produced: that, by publishing 'raw' exam results, the tables measure the quality of the schools' intake rather than of their teaching (for a discussion of 'value-added' measures see Gray *et al.* (2001)); and that, as wealthier and more motivated parents make every effort to ensure their children attend the 'better' schools, there will be a widening social divide between the more and the less successful schools (e.g. Bourdieu & Passeron 1992; Bowe *et al.* 1994; Reay 1998; such claims are, however, disputed by Gorard (2000)).

Owing to the lack of a control group (a problem ubiquitous throughout the social sciences), it would be extremely difficult to investigate the effect of league tables on school standards and social make-up using observations taken 'from life'. (The abolition of league tables in Northern Ire-

* Author for correspondence (r.hoyle@surrey.ac.uk).

land and Wales now provides the possibility of investigating a group where such tables are not published.) Thus, in order to isolate these effects, we have constructed a simple mathematical model that we believe adequately describes the most important factors influencing pupils' exam performance and the exercising of parental choice based on published league tables. In particular, we have investigated both the accuracy of the tables in reflecting school standards and the change in school performance and social segregation over time that result from such choices.

Our most significant observation is that the random year-to-year variations, arising from both the pupil intake and fluctuations in school effects, lead to a poor correlation between the league-table positions and the underlying 'school effect'; surprisingly, this did not improve when we adopted a 'value-added' approach. Nevertheless, some schools still seem to excel, even when we remove all underlying differences between schools. A different method of drawing up the tables, in which we average the exam results over a number of years (the length of a school career), produces much more accurate correlations between the league tables and the underlying school effectiveness.

2. THE MATHEMATICAL MODEL

In order to investigate the effect of league tables on school numbers and school standards, we need to model both the performance of pupils and the movement of families.

We consider all the individual pupils in the exam year group (we have chosen the length of a 'school career' in our model to be 4 years, as an average of schools that run from ages 11 to 16, those that run from 12 to 16 and those that run from 13 to 16) in a relatively isolated unit (e.g. a county), and calculate the league tables for that year based on the average exam performance at each school. In the next year we fill the schools with new pupils from similar social backgrounds, having allowed for some movement of families in reaction to the league tables.

(a) *Pupil performance*

Since league tables are derived from exam results, our aim is to produce a representative 'exam result' for each pupil based on: (i) previous attainment of the pupil; (ii) the influence of the pupil's parents; and (iii) qualities that vary from school to school. Our first task is to try to quantify the relative effect of these different factors.

In the 1970s, Jencks *et al.* (1970) concluded that at least half the variability in pupils' performance could be attributed to differences in their prior attainment and social background (see, for example, Jencks *et al.* 1972). Relatively little of the remaining variation can be directly attributed to the school itself: early studies (Aitkin & Longford 1986) indicated a variation of only 2% directly attributable to school factors, although more recent studies indicate a somewhat higher level of influence (e.g. *ca.* 10% in Willms (1987); 5–8% in Gray & Wilcox (1995, ch. 6); see also Reynolds (1990) and Gorard (2000, pp. 81–82)).

We will assign up to 10% of the variance in pupils' marks to the 'value added' by each school, and consider three possibilities: no difference between schools; a fixed effect for each school accounting for 10% of the total vari-

ation in pupils' marks; and a fixed effect for each school accounting for 5% of the total variation in pupils' marks, along with a random fluctuation each year accounting for another 5% of the variation in marks. While the last of these seems to be the most realistic (see Goldstein 1987, ch. 4), the assumption that the schools provide a 'uniform level of service' is also useful in the context of this study to try to isolate the effect of parental choice on school performance.

Within each school we have chosen to make the 'value added' the same for every pupil. Of course, one would expect there to be a differential effect between pupils according to their ability (among other factors); however, evidence suggests that this is not marked (see Gray & Wilcox 1995, ch. 6).

The influence of parental background on pupils' academic success is now well attested. Indeed, it appears (Gray & Wilcox 1995, ch. 5) that the more detailed the information collected about individual pupils, the greater is the evidence for the importance of a pupil's parents to their school performance. Research by the Nottinghamshire Local Education Authority that took into account the number of professionals in the pupil's household confirms the importance of parental background in academic achievement (see Gray & Wilcox 1995, p. 94).

We have chosen to represent prior attainment and parental factors by one variable each, and have allowed for the possibility of some correlation between the two. In line with observations in Plewis & Goldstein (1998), that prior attainment is responsible for roughly twice as much variation in exam results as socio-economic factors, we have weighted prior attainment twice as heavily as socio-economic factors in determining the final exam score.

All of our pupils are assigned a prior attainment 'score', p , and a number reflecting their socio-economic background, s ; these were chosen independently with a $N(0, 1)$ distribution. (We also performed some simulations in which these two factors were correlated, by taking them from a bivariate normal distribution; we discuss the effect of this change at the end of § 3.) Meanwhile, each school was given a 'value-added' factor that depended on our choice of model: for no school effect we set this factor to zero for every school; for a fixed school effect we assigned each school a number from a $N(0, 1)$ distribution; and when allowing for random fluctuations, we adjusted this each year by taking the average of the underlying effect and an additional $N(0, 1)$ random variable.

The exam result of each pupil is linearly dependent on these three factors. If the school 'value-added' effect is v , the exam result of the pupil is given by

$$C_1(\sqrt{6}p + \sqrt{3}s + v) + C_2.$$

The choice of the constants $\sqrt{6}$ and $\sqrt{3}$ is consistent with the relative effects of the various factors as discussed above. C_1 and C_2 play no part in the calculations and are simply chosen to produce an exam mark in line with the average GCSE score (with $A^* = 8$, $A = 7$, etc.): *ca.* 33, with *ca.* 50% lying between 38 and 28 (cf. observations in Department for Education and Skills 2000).

(b) *League tables*

We used three methods for drawing up the league tables. In the first, which is closest to current practice, we

Figure 1. Schools providing a uniform level of service. (a) Average exam results, (b) number of pupils and (c) average socio-economic level. Each line on the graphs represents a different school. The different line styles (solid, dashed, dotted) are intended to distinguish between schools. The line style for an individual school is consistent between figure parts.

averaged the raw exam results over all pupils in each school, and then ordered the schools according to this figure. In the second, we took the mean of each school's average exam result over a 4-year period, and calculated our league tables on the basis of this revised figure. We would naturally expect these 'averaged' league tables to give a more accurate indication of the underlying 'school effect', because the effect of such averaging is to reduce the amount of variation (e.g. the average of four random variables with a $N(0, 1)$ distribution has the same mean, but a variance of $1/4$). Note that these 'averaged' tables may well yield a better model of parental choice, since one would suspect it to be unusual to base choice of school on the exam results of only the most recent year.

It is not entirely clear, given the simplicity of our model, how to produce a realistic version of a 'value-added' table. If we adjust for prior attainment and socio-economic background, then all we have left is the value-added effect for that particular year. So our value-added tables are, in fact, tables of the school performance for *any given year*.

(c) Exercise of parental choice

The default in our model is to attend the local school, and we anticipate that, despite the current possibility of exercising some limited choice, the fact that preference is still given to those living within a school's catchment area will make moving house the dominant way in which 'parental choice' is exercised.

The parents of those who are to attend school in the following year can consider choosing a school outside their

natural catchment area: in our model families in higher socio-economic groups are more likely to exercise this right (at the very extremes of the distribution, ± 4 , the probabilities of choosing are 0 and 1, and linear between). This takes into account not only the possibility that social factors influence parents' perception of the importance of education, as suggested by various researchers (e.g. Moore & Davenport 1990; Echols *et al.* 1990), but also the various financial factors that have a bearing on how effectively such a choice can be exercised, including moving into the catchment area of a particular school (the 'selection by mortgage' of Hirsch (1997)), the cost of transport to more distant schools (cf. Worpole 1999) and something as prosaic as buying the broadsheet papers that publish sets of league tables.

In our model each family that makes a choice picks the school for their child solely on the basis of the league-table results of the previous year. Note that when the league tables are based on averaged results parents' choices automatically take into account the exam performance over a number of years.

(d) Allocation of school places

Once parental choices have been made (or not), school places are assigned by our computational 'local authority' in a way that favours those living in the catchment area.

Each school has a fixed maximum size, and before a new school year each works out how many spare places it has to offer based on the number of places that were filled

Figure 2. Value-added tables used to simulate fixed school effects. (a) Average exam results, (b) number of pupils and (c) average socio-economic level. Each line on the graph represents a different school. The different line styles (solid, dashed, dotted) are intended to distinguish between schools. The line style for an individual school is consistent between figure parts.

in the previous year; we also impose a maximum increase in the size of any school between one year and the next.

Those families that wish to choose then move into the catchment area of the best school that has places available (these families are taken in a random order to prevent any unfair advantage).

New vacancies may now be available, and a second round gives another chance to those families that were interested in moving but were unable to find a better school with spare places in the first round. Those who did not want to move—or who have been frustrated in both rounds—remain within their catchment area.

At the end of the school year, some schools may have fallen below the minimum numbers viable to keep the school going; in which case the school will be closed. (We give precise details in the next paragraph.)

3. NUMERICAL EXPERIMENTS

We model a situation where there are 10 schools in a neighbourhood and a total of 1000 pupils in each year attending them. At the beginning of the simulation, each school has exactly 100 pupils in each year, and, allowing for a little variation in the numbers, we assign a maximum capacity of 150 pupils per year to each school; numbers are only allowed to increase by 10 each year, and, provided that this creates enough spare places, those schools whose numbers drop below 50 in any one year are closed.

Each pupil attends school for 4 years; thus, it is the exam performance of the pupils who joined the school

4 years ago that affects the league-table positions at the end of the current year. Once a child leaves the school it is replaced by a child from a similar socio-economic background to those of the current first years, such as a younger sibling or a child from the same neighbourhood.

All our figures are plotted from the first year in which parental choice has been exercised (there is a run-in time that allows the first batch of pupils to complete their school career), which is why at the beginning of the plot schools no longer have 100 pupils each.

(a) *No difference between schools*

In our first simulations there are no intrinsic variations between the schools, which allows us to consider in isolation the effect of league-table-driven choice.

With parents basing their decisions solely on the league table from the previous year, one or two schools were quickly perceived as 'failing' and generally had to close within the first 10 years (see figure 1). Within a similar length of time, a small number of schools established themselves as 'successful', and showed an increase in the number of parents in higher socio-economic groups. Such schools generally produced higher exam results, thus maintaining their position. Despite small variations in the socio-economic banding of the schools over time, this was remarkably stable.

Although in many of our simulations the *spread* of the socio-economic indicator is what one would expect from choosing them randomly according to our distribution (the average over 100 pupils has standard deviation 0.1),

Figure 3. Value-added tables used to simulate time-varying school effects. (a) Average exam results, (b) number of pupils and (c) average socio-economic level. Each line on the graph represents a different school. The different line styles (solid, dashed, dotted) are intended to distinguish between schools. The line style for an individual school is consistent between figure parts.

the fact that the indicators for individual schools remain approximately constant over time is more surprising.

So, the very process of choice based on exam-derived league tables can generate a perceived difference in school performance and an ingrained social divide between institutions.

(b) *Value-added tables: an exact reflection of school performance*

At the other extreme are our simulations using 'value-added' tables. Here, the schools do have an effect on the pupils' exam results, and, as discussed in § 2b, the most sensible way to produce value-added tables in our model is to base the tables on the true school effect *in each year*. As one would expect (see figure 2), without random fluctuations from year to year the fact that the league tables present an entirely accurate view of the schools means that they rapidly distinguish between the more successful and the less successful, with the best schools filling quickly and the worst schools closing. A fixed social make-up is established at each school within 7 or 8 years.

(We note a slight peculiarity of our model which means that if a school is full and consistently at the top of the league tables, then its socio-economic make-up remains constant, since no-one will wish to move and there are no extra places available.)

However, in more realistic calculations, where the school effect varies between years, the value-added tables present an inaccurate picture of the relative merits of the different schools: over ten runs, the average rank corre-

lation between the league-table position and the true ranking of the school's underlying effect was only 0.58. It can be seen from the simulations (e.g. figure 3) that there is no clean distinction between the schools, and that the 'market forces' that so quickly stabilized things before lead, in this case, to only a rough match between the worth of a school and its 'market value'. It appears that year-by-year value-added tables may present a very distorted picture of school standards.

However, we also used tables based on the average value added over a period of 4 years. As one would expect, this averaging process greatly decreased the effect of the between-year variations and produced a much better correlation (an average of 0.79 over ten sample runs); nevertheless, the results appeared qualitatively similar to those in figure 3.

(c) *Standard league tables and school performance*

Between these two extremes lie the simulations in which we allow for school effects at levels consistent with observations and use raw exam scores as the basis for the league tables. We expect these simulations to be the most realistic of those presented here.

There are, primarily, two questions. The first, and perhaps the most telling in the light of the results for 'value-added' tables presented in § 3b, is how good a correlation there is between the league-table positions and the true underlying 'school performance'. The second, in the light of our simulations for identical schools, is how great an

Figure 4. League-table simulation for time-varying school effects. (a) Average exam results, (b) number of pupils and (c) average socio-economic level. Each line style represents a different school. The different line styles (solid, dashed, dotted) are intended to distinguish between schools. The line style for each individual school is consistent between figure parts.

effect these tables have on social differentiation between schools.

A simulation with no variation in the underlying school effects from year to year produced an average correlation of 0.86 over ten sample runs, and these high correlations produced results similar to those in figure 2. However, when variation between years was included the figures were not so impressive: the average correlation over ten runs was 0.59. While there was some social banding of schools, there was still some flexibility (see figure 4; the colour plot helps to distinguish between the different schools in this, our most realistic simulation).

Nevertheless, it is noteworthy, and perhaps surprising, that the league tables in this case give no worse a correlation than the value-added tables (in fact our average correlation was marginally greater). Presumably this is because, over time, the better schools attract the children of better-off parents, thus preferentially boosting their exam scores and exaggerating the difference between schools in a way that is reasonably robust to variations between the years.

As we might expect, tables based on averaged exam results over a 4-year period prove much more accurate (an average correlation of 0.75 over ten runs, only moderately lower than the average value-added tables). As ever, with this increased accuracy comes a more pronounced socio-economic banding of the schools.

Finally, we note that introducing a correlation between the prior attainment and the parents' socio-economic group (so that they are chosen from a bivariate normal

distribution) produces qualitatively similar results; although there is a potential for this to promote a social differentiation of schools, such an effect was not pronounced in our simulations, for which we used correlations of 0.2 and 0.4.

4. CONCLUSIONS

We have seen, under the assumptions of our model, that the use of league tables has the potential to cause an entrenched social division between schools (even when all the schools are, in fact, identical). While this is not a desirable outcome, it is difficult to see how to ameliorate this effect while still providing freely available information on schools.

Such social division is more likely to occur when the league tables offer a consistently accurate reflection of the relative ranking of schools. Counter-intuitively, in what we believe to be our most accurate model (with the underlying school effect varying from year to year as well as between schools), the league tables were no less accurate when calculated using the raw exam results than when using a value-added method.

We found throughout that using an exam result averaged over a 4-year period produced a more accurate reflection of underlying school performance, because it lessens the effects of random fluctuations between years. Although such a method would disadvantage a failing school that has been turned around, we suggest that this

would be a more realistic way of assessing school performance than a simple yearly exam report.

Thanks to S. Neill of the Institute of Education, University of Warwick for several interesting conversations, and to Mrs Young at the Warwickshire County Education Offices. Thanks also to the anonymous referees for their helpful suggestions.

REFERENCES

- Aitkin, M. & Longford, N. 1986 Statistical issues in school effectiveness. *J. R. Stat. Soc. A* **149**, 1–42.
- Bourdieu, P. & Passeron, C. 1992 *Reproduction in education, society, and culture*. London: Sage.
- Bowe, R., Gerwitz, S. & Ball, S. 1994 Captured by the discourse? Issues and concerns in researching 'parental choice' *Br. J. Sociol. Educ.* **15**, 63–78.
- Department for Education 1992 *School performance tables, public examination results 1992*. London: HMSO.
- Department for Education and Employment 1996 *Primary school performance tables 1996, Key Stage 2 results*. London: HMSO.
- Department for Education and Skills 2000 *GCSE/GNVQ additional benchmark information*. London: HMSO.
- Department of Health 2001 *NHS performance ratings: acute trusts 2001/01*. London: HMSO.
- Echols, F., McPherson, A. & Willms, J. D. 1990 Parental choice in Scotland. *J. Educ. Policy* **5**, 207–222.
- Education Week 2001 Quality counts 2001: a better balance. *Education Week* **20**(17) (see <http://www.edweek.org/sreports/qc01/>).
- Goldstein, H. 1987 *Multilevel models in educational and social research*. London: Charles Griffin & Co.
- Gorard, S. 2000 *Education and society justice*. Cardiff: University of Wales Press.
- Gray, J. & Wilcox, B. 1995 *Good school, bad school*. Milton Keynes: Open University Press.
- Gray, J., Goldstein, H. & Thomas, S. 2001 Predicting the future: the role of past performance in determining trends in institutional effectiveness at A level. *Br. Educ. Res. J.* **27**, 391–405.
- Hirsch, D. 1997 What can Britain learn from abroad? In *Choice and diversity in schooling: perspectives and prospects* (ed. R. Glatter, P. Woods & C. Bagley), pp. 152–165. London: Routledge.
- Jencks, C., Smith, M., Acland, H., Bane, M. J., Cohen, D., Gintis, H., Heyns, B. & Michelson, S. 1972 *Inequality: a reassessment of the effects of family and schooling in America*. New York: Basic Books.
- Moore, D. & Davenport, S. 1990 Choice: the new improved sorting machine. In *Choice in education: potential and problems* (ed. W. L. Boyd & H. J. Halberg), pp. 187–223. Berkeley, CA: McCutchan.
- Newbold, D. 2000 Italians follow UK lead on testing. *Times Educ. Suppl.*, 20 October 2000, p. 14.
- NHS Executive 1999a *Quality and performance in the NHS: high level indicators*. London: HMSO.
- NHS Executive 1999b *Quality and performance in the NHS: clinical indicators*. London: HMSO.
- Plewis, I. & Goldstein, H. 1998 Excellence in schools: a failure of standards. *Br. J. Curriculum Assessment* **8**, 17–20.
- Reay, D. 1998 Engendering social reproduction: mothers in the educational marketplace. *Br. J. Sociol. Educ.* **19**, 195–209.
- Reynolds, D. 1990 School effectiveness and school improvement: a review of the British literature. In *Judging standards and effectiveness in education* (ed. B. Moon, J. Isaac & J. Powney), pp. 152–169. London: Hodder & Stoughton.
- UK Government Department of Education and Science 1988 *Education Reform Act 1988*. London: HMSO.
- Willms, J. D. 1987 Differences between Scottish educational authorities in their examination attainment. *Oxf. Rev. Educ.* **13**, 211–237.
- Worpole, K. 1999 Driving forces. *Guardian*, 8 June 1999, p. 17.

As this paper exceeds the maximum length normally permitted, the authors have agreed to contribute to production costs.